Ch.2-Bivariate Data

2-variable

Analyzing 2 Categorical Variables

- **You can create categorical variables...** from quant. by grouping quant. responses
- **Categorical Variables are often shown ...** 2 way table

	Frosh	Soph	Jr	Senior	Total
Male	#	#	#		#
Female	#	#			
Total	#				n

Identify:

- **row variable** Gender
- **column variable** Grade/Yr. in school
- **values of the variables** rows/columns
- **overall**
- **total (n)**
- **cells** – where actual data is (#)
- **totals** – summaries

Example:

	Hospital A	Hostpital B	total
Died	63	16	79
Survived	2037	784	2821
total	2100	800	2900

row var = mortality
survival
patient status

Values
died
survived

column var = hospital choice

Values
A
B

$n = 2900$

cells = 4

$P(A) = \frac{2100}{2900}$

$P(D) = \frac{79}{2900}$

2 types of Distributions

1) Marginal Distributions

- How to make: margins (totals) ÷ n (overall total)
- Looking for ... overall % of each value of variables
- ALWAYS in %
- Hospital Example:

Example:

	Hospital A	Hostpital B	total
Died	64	15	79
Survived	2037	784	2821
	2100	800	2900

- Marginal Distribution for ***Survival*** Variable:

Died : 79/2900 = 2.72%

Survived: 2821/2900 = 97.28%

100%

- ***Marginal Distribution for Hospital* Variable:**

A: 2100/2900 = 72.4%

B: 800/2900 = 27.6%

Example:

	Hospital A	Hostpital B	
Died	[illegible]	[illegible]	79
Survived	[illegible]	784	2821
	2100	800	2900

Visually: Bar graphs

1) ***Conditional Distributions***

- ***Look at ...*** one row/column (var.)
- ***Then look at ...*** one value of the var.
- ***Break down ...*** that value into its parts
- ***ALWAYS*** in %
- ***Hospital Example:***
  - ***Conditional Distribution for Survival Variable:***

Example:

	Hospital A	Hostpital B	
* Died	63	16	79
* Survived	2037	784	2821
	2100	800	2900

Died

A: $63/79 = 79.7\%$

B: $16/79 = 20.3\%$

100%

Survived

A: $2037/2821 = 72.2\%$

B: $784/2821 = 27.8\%$

100%

- ***Conditional Distribution for Hospital* Variable:**

A
D: 63/2100 = 3%
S: 2037/2100 = 97%

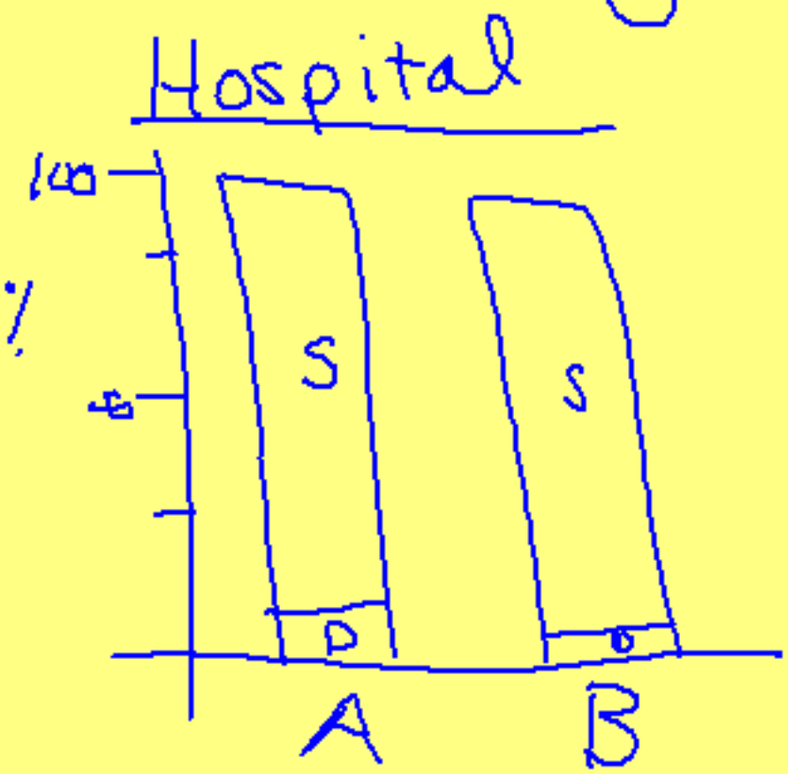
B
D: 16/800 = 2%
S: 784/800 = 98%

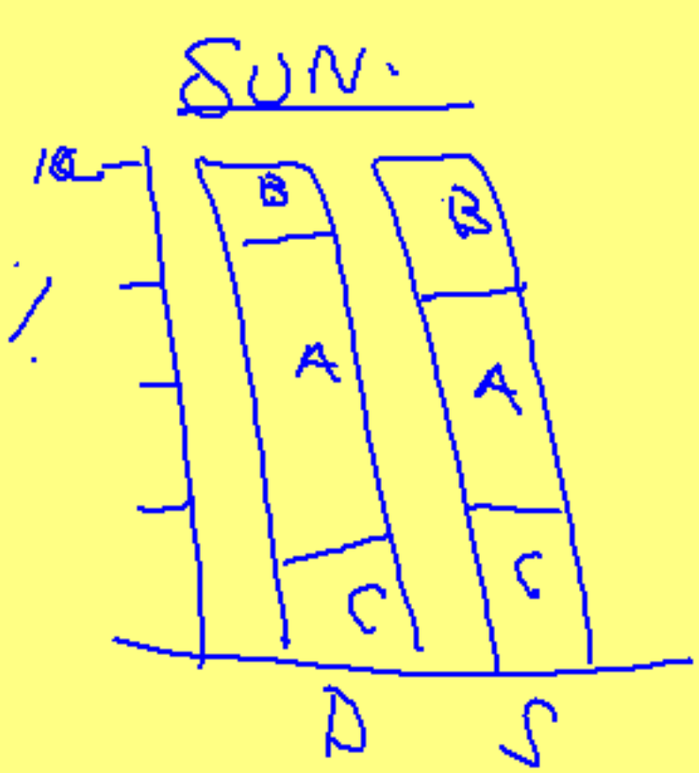
Example:

	Hospital A	Hostpital B	
Died	63	16	79
Survived	2037	784	2821
	2100	800	2900

Represented Visually: segmented/stacked bar graph

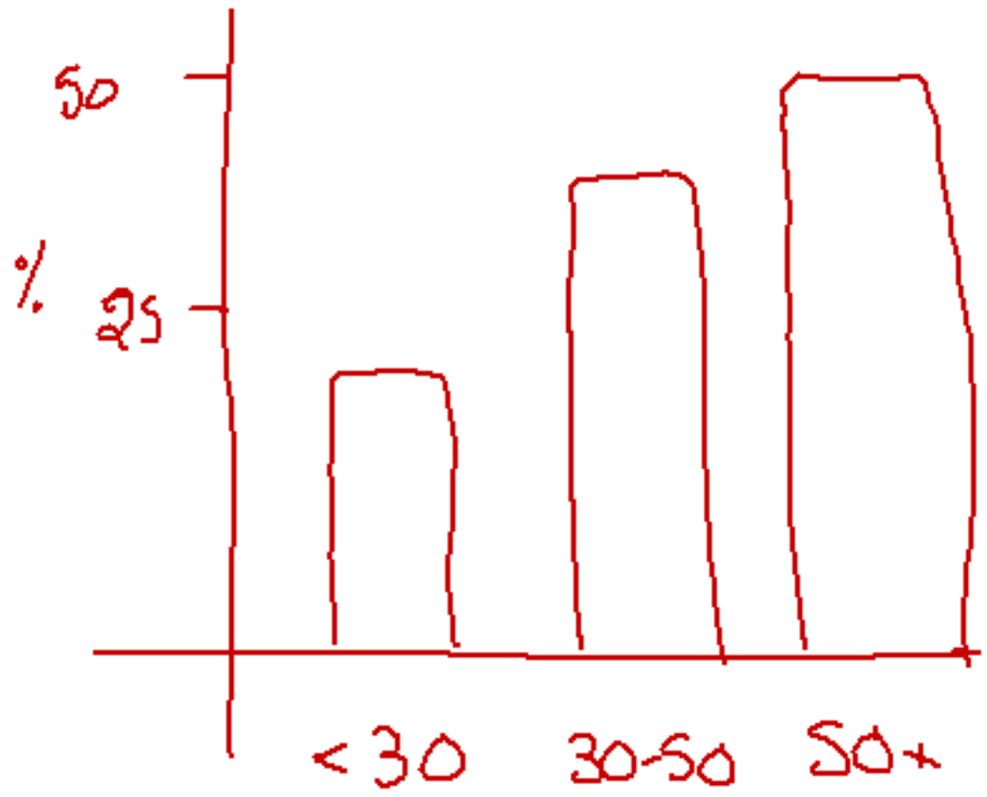
- **Each bar = 100%**
- **Conditional Variable on ...** x-axis
- **Bars are ...** segmented/split into diff. values





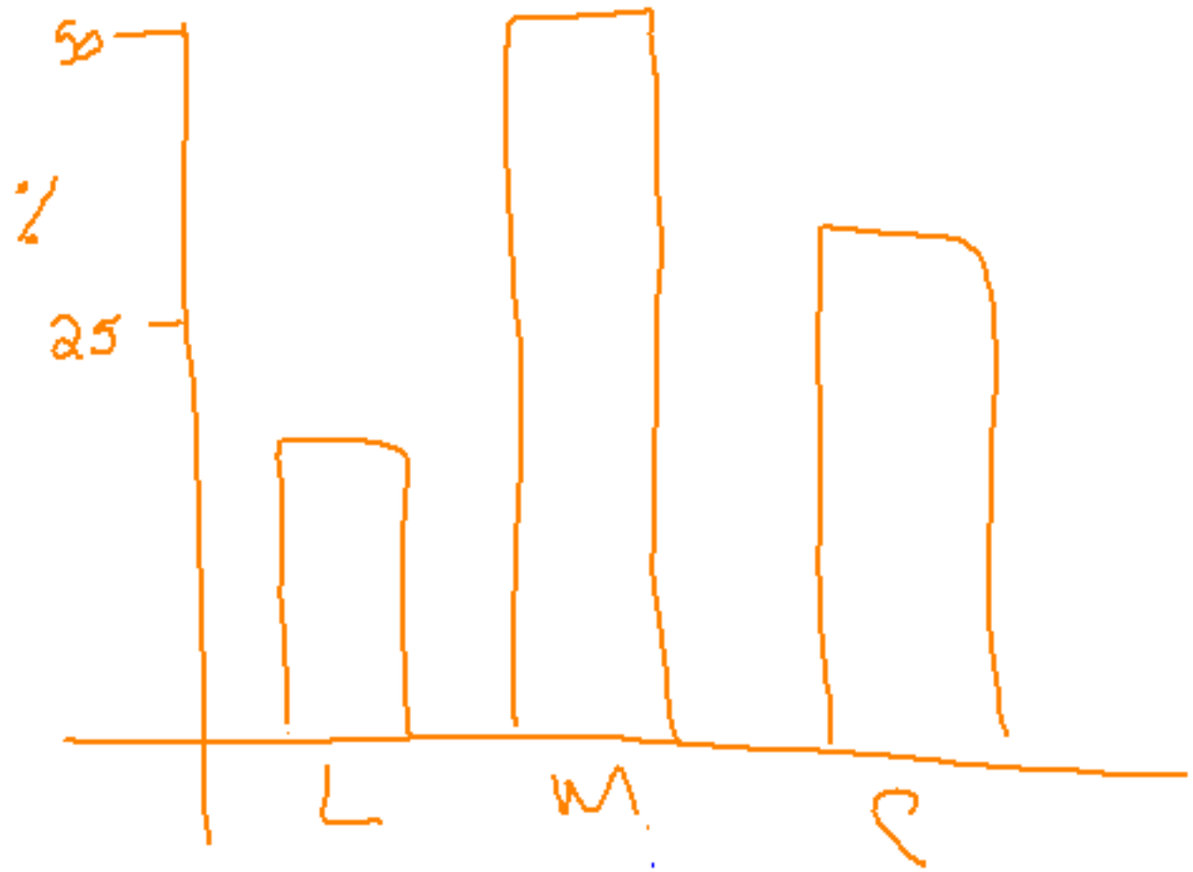
Marginal

Age



Age

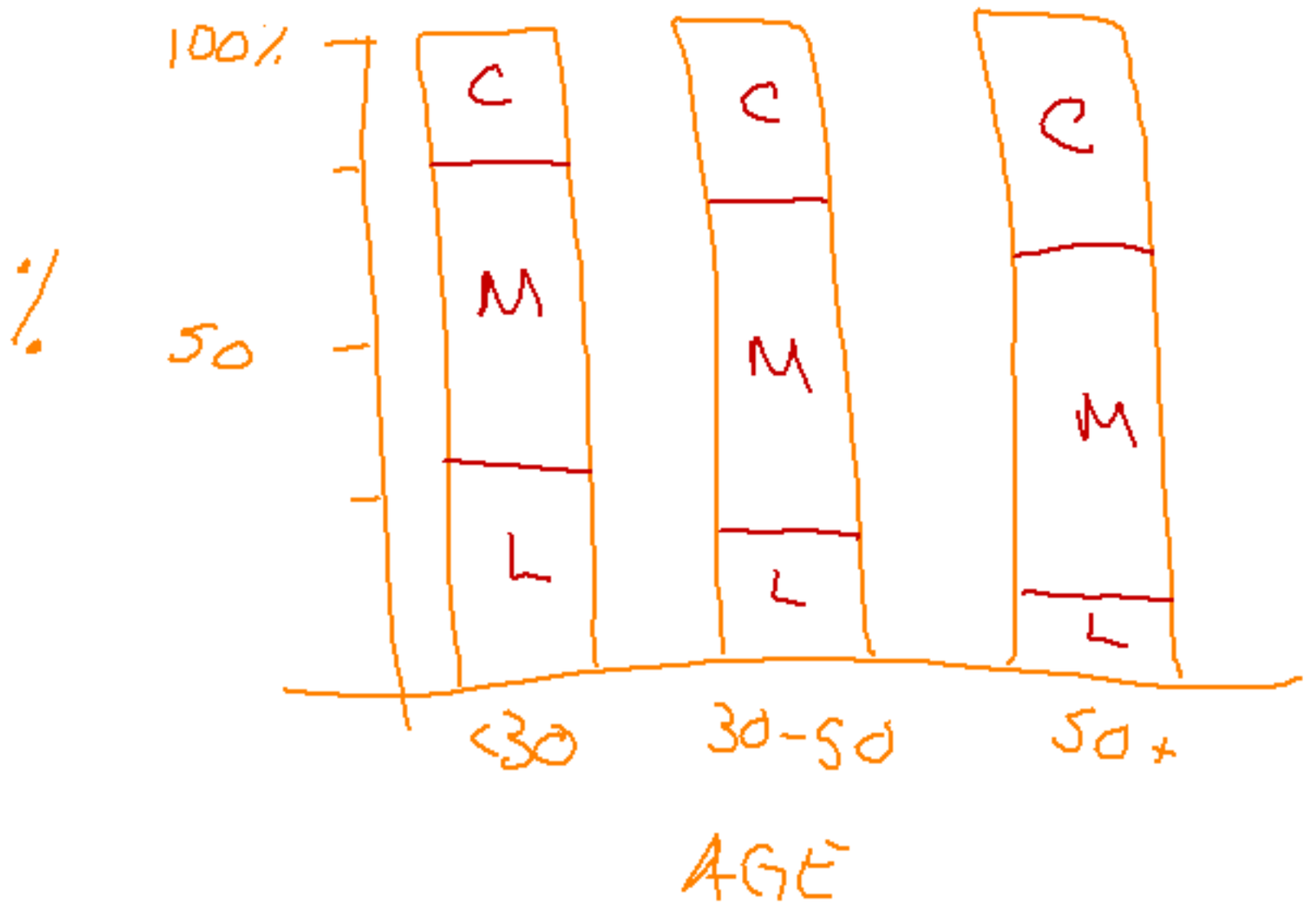
Political Pref.



Ideology

Conditional - Age

	<30	30-50	50+
L	28%	21.3%	15%
M	47.3%	50%	48.5%
C	24.7%	28.8%	36.5%



Independence

Simpson's paradox